Protein Structure

Why protein folds?
Secondary structure
  Alpha-helix (\(\alpha\)-helix)
  Beta-sheet (beta-conformation)
  Beta-turn (\(\beta\)-turn)
  Ramanchandran plot
Tertiary & Quaternary Structure
Motif & Domain
Stable conformation
Visualization of protein
Denature of protein

Much More About Structure

Structure ↔ Function
Structure ↔ Mechanism
Structure ↔ Origins/Evolution
Structure-Based Drug Design
Solving the Protein Folding Problem
Protein Conformation

Native protein

Why Protein Folds?

Peptide bond
Disulfide bond
Hydrogen bonding
Ionic interaction
Hydrophobic-lipophilic interaction
Van der Waals interaction
Rotational prohibition
Figure 4-2b
Lehrerker Principles of Biochemistry, Fifth Edition
© 2006 W.H. Freeman and Company

Figure 4-7b
Lehrerker Principles of Biochemistry, Fifth Edition
© 2006 W.H. Freeman and Company
Figure 4.26: Native state, catalytically active state. Disulfide cross-links correctly re-formed.
Levels of Protein Structure

- **Primary**
  - Linear AA sequence
  - Covalent bonds

- **Secondary**
  - Local structure; certain “motifs” are common
  - Mostly H-bonds

- **Tertiary**
  - Complete 3D shape
  - H-bonds, hydrophobic interactions, ionic bonds, van der Waals interactions, disulfide bonds

- **Quaternary**
  - >1 peptide chain
  - Mostly H-bonds

---

Protein Folding

- **Folded shape = conformation**
- **Three-dimensional, functional structure = native**
  - Energy of native conformation?

- **Molecular chaperones**
- There are thousands of possible conformations, but not an infinite amount...

- **Conformations are restrained by**
  - Planarity of peptide bond
  - “allowed” angles

- **No algorithm predicts the 3D shape with high accuracy**
Primary Structure

• AA sequence of polypeptide chain(s)
• Linked by peptide bonds
• Linear sequence
• Predication of primary structure?
• Experimental determination: protein sequencing

Secondary Structure

• Regular repeating structure
  – Helices
  – Sheets
• Torsion/dihedral angles
  – Angles of rotation around $C_i$
  – Clockwise (+) and counterclockwise (-)
  – $\Phi$ = rotation around $C_i$-N
  – $\Psi$ = rotation around $C_i$-C
• How free is rotation?
  – Not very (sterics)
  – Avoid collision of C=O, N-H, R
  – Calculations of allowed values = Ramachandran diagram
Alpha Helix

- (30-35%)
- $\Phi = -57^\circ$, $\Psi = -47^\circ$
- Discovered by Pauling: 1951
- $\alpha$-helix formers: A,C,L,M,E,Q,H,K
- Tightly wound, repeating sequence
- “Right-handed”
- Each twist $\sim 5.4$ Å; 3.6 residues
- Average length = 18 residues
- R-groups are on outside of helix
- Stabilized by H-bonds between C=O (i) and N-H (i + 4)

Alpha helix, cont.

- Deviates from ideal conformation at ends (less H-bonding)
- Some amino acids are “$\alpha$-helix breakers”
  - Repeating like-charges
  - Repeating “bulky” groups
  - Pro and Gly
- Effects on helical stability:
  - Electrostatic interactions between adjacent residues
  - Steric interference between adjacent residues
  - Interactions between residues 3-4 amino acids away
  - Polarity of residues at both ends of helix (positive at amino end; negative at carboxyl)
### TABLE 4-1  Propensity of Amino Acids to Take Up an α-Helical Conformation

<table>
<thead>
<tr>
<th>Amino acid</th>
<th>( \Delta \Delta G^\circ ) (kJ/mol)*</th>
<th>Amino acid</th>
<th>( \Delta \Delta G^\circ ) (kJ/mol)*</th>
</tr>
</thead>
<tbody>
<tr>
<td>Ala</td>
<td>0</td>
<td>Leu</td>
<td>0.79</td>
</tr>
<tr>
<td>Arg</td>
<td>0.3</td>
<td>Lys</td>
<td>0.63</td>
</tr>
<tr>
<td>Asn</td>
<td>3</td>
<td>Met</td>
<td>0.88</td>
</tr>
<tr>
<td>Asp</td>
<td>2.5</td>
<td>Phe</td>
<td>2.0</td>
</tr>
<tr>
<td>Cys</td>
<td>3</td>
<td>Pro</td>
<td>&gt;4</td>
</tr>
<tr>
<td>Gln</td>
<td>1.3</td>
<td>Ser</td>
<td>2.2</td>
</tr>
<tr>
<td>Glu</td>
<td>1.4</td>
<td>Thr</td>
<td>2.4</td>
</tr>
<tr>
<td>Gly</td>
<td>4.6</td>
<td>Tyr</td>
<td>2.0</td>
</tr>
<tr>
<td>His</td>
<td>2.6</td>
<td>Trp</td>
<td>2.0</td>
</tr>
<tr>
<td>Ile</td>
<td>1.4</td>
<td>Val</td>
<td>2.1</td>
</tr>
</tbody>
</table>


*\( \Delta \Delta G^\circ \) is the difference in free-energy change, relative to that for alanine, required for the amino acid residue to take up the α-helical conformation. Larger numbers reflect greater difficulty taking up the α-helical structure. Data are a composite derived from multiple experiments and experimental systems.

Table 4-1  Lehninger Principles of Biochemistry, Fifth Edition
© 2008 W.H. Freeman and Company

---

**Figure 4-5**

- Amino terminus: **δ**
- Carboxyl terminus: **δ**

Lehninger Principles of Biochemistry, Fifth Edition
© 2008 W.H. Freeman and Company
**β-Sheet / β-Strand**

- Extended, zigzag conformation (20-25%)
  - Hydrogen bond between groups across strands
  - Forms parallel and antiparallel pleated sheets
  - Residues alternate above and below β-sheet
  - β-sheet formers: V, I, P, T, W
- **Intrastrand H-bonding**
- Average 6 residues/strand; up to 15
- 2-12 strands/sheet; average 6
- R-groups alternate on opposite sides of sheet
- Distortions:
  - Beta-bulge = extra residue
  - Kink = Pro

**Anti-parallel vs. Parallel**

- **Anti-parallel β-sheet**
  - Opposite orientation
  - $\phi = -140^\circ$, $\psi = 135^\circ$
  - More stable
  - Can be twisted
  - 6.5 Å per two amino acid residues
  - Can withstand distortions and exposure to solvent
- **Parallel β-sheet**
  - Same amino-carboxyl direction
  - Less twisted
  - Tend to be buried
  - $\phi = -120^\circ$, $\psi = 115^\circ$
  - 7.0 Å per two amino acid
- Can have mix of parallel and anti-parallel
Antiparallel

Top view

Side view

Parallel

Top view

Side view
**β-turns**

- Interacting strands can be many amino acids apart
- Turns are 180°; “connect” strands in folded (globular) proteins
- Interaction is between carbonyl oxygen of AA 1 and amino hydrogen of AA 4
  - Short turn (4 residues)
  - Hydrogen bond between C=O & NH groups within strand (3 positions apart)
  - Usually polar, found near surface
  - β-turn formers: S, D, N, P, R
- Pro and Gly are often present
  - Gly: small and flexible (Type II turns)
  - Pro: cis conformation makes inclusion in tight turn favorable
Others

• Loop
  – Regions between α-helices and β-sheets
  – On the surface, vary in length and 3D configurations
  – Do not have regular periodic structures
  – Loop formers: small polar residues

• Coil (40-50%)
  – Generally speaking, anything besides α-helix, β-sheet, β-turn
Figure 4-8a
Lelievre: Principles of Biochemistry, Fifth Edition
© 2006 W.H. Freeman and Company

Figure 4-9
Lelievre: Principles of Biochemistry, Fifth Edition
© 2006 W.H. Freeman and Company
Figure 4.17a
Lehrbucher: Principles of Biochemistry, Fifth Edition
© 2006 W. H. Freeman and Company

**β-α-β Loop**
Figure 4-17b
Lehrer's Principles of Biochemistry, Fifth Edition
© 2008 W.H. Freeman and Company

β Barrel

(a) Typical connections in an all-β motif
(b) Right-handed connection between β strands
(c) Twisted β sheet

Figure 4-19
Lehrer's Principles of Biochemistry, Fifth Edition
© 2008 W.H. Freeman and Company

Crossover connection (not observed)
Left-handed connection between β strands (very rare)
Two types of cyclic symmetry

$C_2$  

Twofold

$C_3$  

Threefold

Figure 4-23a
Lehninger Principles of Biochemistry, Fifth Edition
© 2008 W.H. Freeman and Company
Two types of dihedral symmetry

Icosahedral symmetry

Figure 4-23b
Lehninger Principles of Biochemistry, Fifth Edition
© 2008 W.H. Freeman and Company

Figure 4-23c
Lehninger Principles of Biochemistry, Fifth Edition
© 2008 W.H. Freeman and Company
Proteins are Complex

- Average residue contains 8 “heavy” atoms
- Average protein contains 300 amino acids
- Average structure contains 2400 atoms
- First structure (sperm whale myoglobin) took ~5 years with a team of ~15 key punch operators working around the clock to solve
- Most structures still take 1 year to solve

Solving Protein Structures

- Only 2 kinds of techniques allow one to get atomic resolution pictures of macromolecules
- X-ray Crystallography (first applied in 1961 - Kendrew & Perutz)
- NMR Spectroscopy (first applied in 1983 - Ernst & Wuthrich)
X-ray Crystallography

- Crystallization
- Diffraction Apparatus
- Diffraction Principles
- Conversion of Diffraction Data to Electron Density
- Resolution
- Chain Tracing
Classification of Proteins by Secondary Structure

• Fibrous
  – High composition of single secondary structure
  – Strong and flexible
  – Collagen
    • Triple helix (left-handed helices, right-handed super helix)
  – Silk fibroin
    • Anti-parallel β-sheet
  – α-Keratin
    • Left-handed coiled coil

• Globular
  – Majority of all proteins
  – Contain several types of secondary structure (regular and non-regular)
  – Percentage of protein (on average):
    • 31% α-helix
    • 28% β-sheet
    • 13% turns/bends
    • 28% loops and random coil

<table>
<thead>
<tr>
<th>Structure</th>
<th>Characteristics</th>
<th>Examples of occurrence</th>
</tr>
</thead>
<tbody>
<tr>
<td>α Helix, cross-linked by</td>
<td>Tough, insoluble protective structures of varying</td>
<td>α-Keratin of hair, feathers, and nails</td>
</tr>
<tr>
<td>disulfide bonds</td>
<td>hardness and flexibility</td>
<td></td>
</tr>
<tr>
<td>βI Conformation</td>
<td>Soft, flexible filaments</td>
<td>Silk fibroin</td>
</tr>
<tr>
<td>Collagen triple helix</td>
<td>High tensile strength, without stretch</td>
<td>Collagen of tendons, bone matrix</td>
</tr>
</tbody>
</table>

Table 4-2: Secondary Structures and Properties of Some Fibrous Proteins

Table by Lehninger Principles of Biochemistry, 7th Edition
© 2000, W.H. Freeman and Company
Aspects Which Determine Tertiary Structure

- Covalent disulfide bonds from between closely aligned cysteine residues form the unique Amino Acid cystine.
- Nearly all of the polar, hydrophilic R groups are located in the surface, where they may interact with water.
- The nonpolar, hydrophobic R groups are usually located inside the molecule.

Protein Visualization Model

[Images of CPK, Line, Trace, Cartoon representations of a protein structure]
Protein Visualization Model

Cylinder  Ribbon (N-C gradient)

Protein Visualization Model

Ribbon (2° structure)  Stick
Protein Visualization Model

Space Filling  Wire Frame (Vector)

Protein Visualization Model

Wireframe  Ball and stick
### Visualization Software

- Biodesigner
- CACTVS
- Chemdraw net Plugin
- Chemical2vmd
- Chime
- Chimera
- Cn3D
- CONSCRIPT
- Dino
- Flex
- FlexV
- Garlic
- Gdis
- gOpenMol
- GRASP
- Hyperactive Molecules Using Chemical MIME
- ICMLite
- ImageMagick
- INTERCHEM modelling software
- Jmol
- Kinemage
- MacMolecule 2 and PCMolecule 2
- Maestro
- Marvin Applets and JavaBeans
- Mercury
- MindTool
- Molden
- MOLEKEL
- MOLMOL
- MolPov
- MolScript
- MolView and MolView Lite
- MSP
- NIH Image
- O
- ORTEX
- POV-Ray
- Pov4Grasp
- PovChem
- Protein Explorer
- PS88
- PyMOL
- Qmol
- QTree
- RasMol
- Raster3D
- Ribbons
- RNA Movies
- RNA Models
- RNAViz
- Spock
- Swiss-PdbViewer
- Tachyon
- VEGA
- Viewmol
- VMD
- WebLab ViewerLite
- WebMol
- WinMGM
- XMol

### MolScript

- Web Site: [http://www.avatar.se/molscript/](http://www.avatar.se/molscript/)
- MolScript is a program for displaying molecular 3D structures, such as proteins, in both schematic and detailed representations.
- Helps from scripting language, we can render the representation by ourselves.
Molscript Representations of Typical Protein Architectures.

- β 7 Propellor (2bbkH)
- α 4-Layer Sandwich (2dnjA)
- β Barrel (2por)
- Sandwich (2hlaB)

- α β Barrel (4timA)
- α Helix Bundle (2ccy)
- 2 Solenoid (1tsp)
- α Horseshoe (1bnh)
RasMol/OpenRasMol

- RasMol is a program for molecular graphics visualization originally developed by Roger Sayle.
- “Grand-daddy” of all visual freeware
- The most popular one viewer.
Jmol

- Jmol is a free, open source molecule viewer for students, educators, and researchers in chemistry and biochemistry.
- It’s cross-platform, running on Windows, Mac OSX, and Linux/UNIX systems.
- It has more detailed materials on how to write a script language and an interactive web UI to demonstrate.
- It’s a Java based application.

1ATP vs. 1BO1

PDBID: 1ATP | PDBID: 1BO1
PyMOL

- PyMOL is a user-sponsored molecular visualization system on a open-source foundation.
- Web Site: http://pymol.sourceforge.net/
- It’s a Python based application.

DeepView (Swiss-PDB Viewer)

- Swiss-PDB viewer is an application that provides a user friendly interface allowing to analyze several proteins at the same time.
- Swiss-PDB viewer is not just a viewer and more powerful than the other viewers.
- It also supports SWISS-MODEL, a fully automated protein structure homology-modeling server.
- It also provides a function of structure alignment.
The loss of secondary, tertiary, or quaternary protein structure due to disruption of noncovalent interactions and/or disulfide bonds that leaves peptide bonds and primary structure intact.
Factors That Cause Protein to Denature

Heat
pH change
Hydrogen bonding reagents
Detergents
Non-polar solvents
Extremely high or low salt concentration
Oxidants/reductants
Mechanical Agitation

Figure 4-27
 Lehninger Principles of Biochemistry, Fifth Edition
© 2000 W.H. Freeman and Company
What is protein disorder?

Disordered regions (DRs) are entire proteins or regions of proteins which lack a fixed tertiary structure, essentially being partially or fully unfolded. Such disordered regions have been shown to be involved in a variety of functions, including DNA recognition, modulation of specificity/affinity of protein binding, molecular threading, activation by cleavage, and control of protein lifetimes. Although these DRs lack a defined 3-D structure in their native states, they frequently undergo disorder-to-order transitions upon binding to their partners.

Depiction of degrees of protein disorder:
- 0 – totally ordered
- 1-5 – partial disorder
- 6,7 – total disorder

- **Heat:** The weak side-chain attractions in globular proteins are easily disrupted by heating, in many cases only to temperatures above 50°C.

- **Mechanical agitation:** The most familiar example of denaturation by agitation is the foam produced by beating egg whites. Denaturation of proteins at the surface of the air bubbles stiffens the protein and causes the bubbles to be held in place.

- **Detergents:** Even very low concentrations of detergents can cause denaturation by disrupting the association of hydrophobic side chains.
- **Organic compounds**: Organic solvents can interfere with hydrogen bonding or hydrophobic interactions. The disinfectant action of ethanol results from its ability to denature bacterial protein.

- **pH change**: Excess H⁺ or OH⁻ ions react with the basic or acidic side chains in amino acid residues and disrupt salt bridges. An example of denaturation by pH change is the protein coagulation that occurs when milk turns sour because it has become acidic.

- **Inorganic salts**: Sufficiently high concentrations of ions can disturb salt bridges.

- Most denaturation is irreversible. Hard-boiled eggs do not soften when their temperature is lowered.
Figure 4.38
Lehninger Principles of Biochemistry, Fifth Edition
© 2000 W.W. Freeman and Company

<table>
<thead>
<tr>
<th>kcal/mol</th>
<th>kJ/mol</th>
<th>lnK^eq</th>
<th>K^eq</th>
</tr>
</thead>
<tbody>
<tr>
<td>1</td>
<td>4.184</td>
<td>1.69</td>
<td>5.41</td>
</tr>
<tr>
<td>2</td>
<td>8.368</td>
<td>3.38</td>
<td>29.23</td>
</tr>
<tr>
<td>5</td>
<td>20.92</td>
<td>8.44</td>
<td>4619.71</td>
</tr>
<tr>
<td>10</td>
<td>41.84</td>
<td>16.88</td>
<td>2134179.78</td>
</tr>
<tr>
<td>15</td>
<td>62.76</td>
<td>25.31</td>
<td>98592623925.87</td>
</tr>
<tr>
<td>20</td>
<td>83.68</td>
<td>33.75</td>
<td>455469429828309.00</td>
</tr>
<tr>
<td>25</td>
<td>104.6</td>
<td>42.19</td>
<td>2104137137724550000.00</td>
</tr>
<tr>
<td>30</td>
<td>125.52</td>
<td>50.63</td>
<td>9720505492587280000000.00</td>
</tr>
<tr>
<td>35</td>
<td>146.44</td>
<td>59.07</td>
<td>4490592620479100000000000.00</td>
</tr>
<tr>
<td>40</td>
<td>167.36</td>
<td>67.50</td>
<td>2074524015081250000000000000000.00</td>
</tr>
<tr>
<td>45</td>
<td>188.28</td>
<td>75.94</td>
<td>958370142399971000000000000000000000000.00</td>
</tr>
<tr>
<td>50</td>
<td>209.2</td>
<td>84.38</td>
<td>44273930943516700000000000000000000000000000.00</td>
</tr>
<tr>
<td>55</td>
<td>230.12</td>
<td>92.82</td>
<td>2045327660440840000000000000000000000000000000000000.00</td>
</tr>
<tr>
<td>60</td>
<td>251.04</td>
<td>101.26</td>
<td>9448822703141950000000000000000000000000000000000000000000.00</td>
</tr>
</tbody>
</table>